

# Gábor Szabó

gabor.j.szabo@gmail.com

+1-574-386-4195

1121 Broadway  
Belmont, CA 94002

## PROFESSIONAL EXPERIENCE

---

Nov 2016– **Tesla Autopilot, Palo Alto, CA**

Aug 2018 Senior Staff Software Engineer & Team Lead

- *Tesla Autopilot Mapping*. Led the Mapping team for Autopilot. My team developed the backend processing algorithms and infrastructure, map representations, and firmware components for using high-resolution maps in Tesla's Autopilot.
- *Maps for localization and control*. My team designed & implemented a new end-to-end mapping pipeline to deliver high-volume map data to customers' cars securely. Worked with the Localization, Controls, Vision, and Cloud teams to efficiently integrate the new map features in firmware.
- *Building high-accuracy maps from sensor data*. Led the creation of new HD map objects from telemetry streams, used both for on-board control and backend vision training. Defined the roadmap, integration strategies, and technologies and implementations to be used.
- *Individual contributions*. Implemented routing algorithms for navigation, and map synchronization and multiprocessing frameworks in firmware. Unblocked map building pipeline by rewriting the previous process as a distributed service. Mentored and built team, executed firmware planning, triage, and deployment processes.

July 2015– **Lyft, Inc., San Francisco, CA**

Oct 2016 Senior Data Scientist

- *Predicting estimated times of arrival*. Improved travel time estimations used in driver dispatch and matching, resulting in 10–15% of gains in prediction accuracies.
- *Demand forecasting*. Led project to develop a demand forecasting framework to predict ride requests in real-time. Achieved significant improvements over the existing method.
- *Driver repositioning*. Initiated project to re-balance driver supply according to predicted future ride demand with global optimization, by giving drivers optimal destinations.
- *Conversion modeling*. Modeled the conversion propensities of passengers who open the Lyft application to potentially request a ride.

May 2011– **Twitter, Inc., San Francisco, CA**

June 2015 Staff Data Scientist

- *Community detection of Twitter users*. Implemented an algorithm to find tight communities, scaling to Twitter's social graph. Technical lead of the team for research and productionizing.
- *Event detection in short messages*. Researched and implemented real-time conversation topic and bursty activity detection in Tweets. Built a streaming distributed system with Storm and Mesos.
- *Advertiser modeling and revenue forecasting*. Built models to describe advertiser and user behavior on anomalous days. Created frameworks for revenue forecasting and opportunity sizing.
- *User and content modeling*. Built statistical models for user engagements and recommendations. Tools used include R, Python, Java, and Hadoop frameworks (Cascading, Pig).
- *Tools for distributed data analytics*. Author of PyCascading, an open-source framework to write MapReduce jobs fully in Python. Co-developed an evaluation framework for ad-hoc experiments.

Jul 2007– **Hewlett-Packard Laboratories, Social Computing Laboratory, Palo Alto, CA**

Apr 2011 Research Scientist, working with HP Labs Senior Fellow Bernardo Huberman

- *Predicting the popularity of online content*. Developed a predictive algorithm for the long-term popularity of content on Twitter, Youtube, and Digg, using early rates of access.
- *Friendlee: A mobile application for automatic social ties*. Developed a mobile application to replace the mobile phone book using contact discovery. First prize at the RCS Developer Challenge, showcased at the 2010 Mobile World Congress.
- *Relationship types and diversity in online communities*. Analyzed the social network and long-tail user behavior in the political community site Essembly.

- Sep 2005– **Harvard Medical School, Boston, MA**
- Jun 2006 Postdoctoral Fellow and Research Lead with Orange, working with Prof. Albert-László Barabási
- *Weak and strong links in society*. Analyzed the social network and communications of mobile phone subscribers over extended periods.
  - *Human protein interaction networks*. Analyzed the importance of proteins linked to a human disease in their interaction network, and found that only a few proteins play a central role.
- Nov 2004– **University of Notre Dame, Department of Physics, Notre Dame, IN**
- Nov 2006 Postdoctoral Fellow, working with Prof. Albert-László Barabási
- *Mobile service usage*. Studied word-of-mouth and mass media effects on the diffusion of mobile services. Developed an accurate predictive model, suggested a new paradigm for viral marketing.
  - *Churn prediction*. Advised Orange on new mobile services and promotions. Developed algorithms based on social interactions, significantly improving the accuracy of churn predictions.
  - *Large-scale mobile phone-based alert and prediction system*. Studied how to detect and forecast anomalous behavior of crowds using existing wireless infrastructure.

## TECHNICAL EXPERIENCE

---

### Languages

- Proficient: C/C++ • Java • Python • R • shell scripting
- Used extensively: Perl • x86 assembly • Pascal • OPL

### Libraries, tools

- Hadoop MapReduce • Apache Flink, Spark, Storm, Pig, Mesos • Cascading • MPI
- Boost C++ library • MATLAB • OpenGL • wxWindows (cross-platform GUI API)
- SQL variants • Vertica • MongoDB
- Android (application programming) • Linux • Windows • DOS • Symbian EPOC • Atmel AVR

### Open source

- Author of PyCascading (<https://github.com/twitter/pycascading>), a MapReduce framework in Python.

## EDUCATION

---

- Sep 2001– **Budapest University of Technology and Economics, Hungary**
- Sep 2004 Ph.D. in Statistical and Computational Physics (summa cum laude), advisor: János Kertész
- Sep 1996– **Budapest University of Technology and Economics, Hungary**
- Jun 2001 M.Sc. in Physics, advisor: János Kertész

## SELECTED BOOKS, PUBLICATIONS, PATENTS

---

1. **G. Szabó**, G. Polatkan, O. Boykin, A. Chalkiopoulos, “*Social Media Data Mining and Analytics*”, **Wiley** (2018).
2. V. Muralidhar, **G. Szabó**, “*Topological analysis of biomolecular networks*”, in “*Systems Bioinformatics: An Engineering Case-based Approach*, by G. Alterovitz and M.F. Ramoni”, **Artech House** (2007).
3. **G. Szabó**, B.A. Huberman, “*Predicting the popularity of online content*”, **Communications of the ACM** **53** (8), 80–88 (August 2010).
4. J.-P. Onnela, J. Saramäki, J. Hyvönen, **G. Szabó**, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, “*Structure and tie strengths in mobile communication networks*”, **Proceedings of the National Academy of Sciences** **104**, 7332–7336 (2007).
5. J. Lim, T. Hao, C. Shaw, A.J. Patel, **G. Szabó**, J.-F. Rual, C.J. Fisk, N. Li, A. Smolyar, D.E. Hill, A.-L. Barabási, M. Vidal, H.Y. Zoghbi, “*A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration*”, **Cell** **125**, 801–814 (2006).
6. *Managing online content based on its predicted popularity*, US Patent 8296253 B2 (2009).
7. *Social networking of mobile devices*, US Patent 20110319058 A1 (2009).